Indexing Consistency in MEDLINE*

BY MARK E. FUNK, *Head, Collection Development*

CAROLYN ANNE REID, *On-Line Services Coordinator*
Midcontinental Regional Medical Library Program

Leon S. McGoogan Library of Medicine
University of Nebraska Medical Center
Omaha, Nebraska 68105

ABSTRACT

The quality of indexing of periodicals in a bibliographic data base cannot be measured directly, as there is no one "correct" way to index an item. However, consistency can be used to measure the reliability of indexing. To measure consistency in MEDLINE, 760 twice-indexed articles from 42 periodical issues were identified in the data base, and their indexing compared. Consistency, expressed as a percentage, was measured using Hooper's equation. Overall, checktags had the highest consistency. Medical Subject Headings (MeSH) and subheadings were applied more consistently to central concepts than to peripheral points. When subheadings were added to a main heading, consistency was lowered. "Floating" subheadings were more consistent than were attached subheadings. Indexing consistency was not affected by journal indexing priority, language, or length of the article. Terms from MeSH Tree Structure categories A, B, and D appeared more often than expected in the high-consistency articles; whereas terms from categories E, F, H, and N appeared more often than expected in the low-consistency articles. MEDLINE, with its excellent controlled vocabulary, exemplary quality control, and highly trained indexers, probably represents the state of the art in manually indexed data bases.

UNDERSTANDING how an information-retrieval system performs can aid both searchers and producers of the system. For the producer, performance levels can be used as a basis to improve design. For the searcher, using a system's strengths can improve retrieval effectiveness. Manual indexing, the assignment of subject headings to a document, probably is the most critical aspect in any system's performance. We studied the consistency of manual indexing in MEDLINE.

There is, of course, no one "correct" way to index a document, just as there is no one correct way to write a document. Language, unlike numbers, is imprecise and full of nuance and shades of meaning. Sir Max Beerbohm, writing about the English language, called it "beautifully vague . . . with its meanings merging into one another as softly as the facts of landscape in the moist English climate." Indexers, even when using a controlled vocabulary, face a multitude of difficult choices and decisions. To investigate the problem of assigning subject headings, many studies have measured interindexer consistency, i.e., the agreement of subject headings chosen by different indexers for the same document [1].

Almost all of these studies had assumed that highly consistent indexing was inherently good—that it would result in better searches. This relationship is logically appealing: if indexing is consistent, then searchers can rely on term A to retrieve information on subject A and articles irrelevant to subject A will not be retrieved. Likewise, no additional terms are needed to ensure that subject A is retrieved, which unnecessarily broaden the search results.

However, the relevance of measuring indexing consistency was challenged by Cooper [2]. He argued that because bad indexing could have high levels of consistency, measuring these levels would tell us nothing of a system's performance. In an experiment that seems to have ended this debate, Leonard [3] tested the effect of indexing consistency levels on retrieval effectiveness, and concluded that "interindexer consistency and retrieval effectiveness exhibit a tendency toward a direct, positive relationship, i.e., high interindexer consistency in assignment of terms appears to be associated with a high retrieval effectiveness of the documents indexed. The concern for maintaining high consistency in indexing seems to have been valid and the work of developing and applying indexing rules and vocabulary controls seems justified." Thus, there seems to be ample reason for measuring this aspect of indexing.

*Based on a paper presented at the Eighty-first Annual Meeting of the Medical Library Association, Montreal, Quebec, Canada, 1981.

Previous MEDLINE Studies

Very few studies of indexing consistency have involved the MEDLARS/MEDLINE data base. Fortunately, the few that do all used the same mathematical method of calculating consistency, so their results can be compared. The method, developed by Hooper [4], is shown in Figure 1. With this equation, consistency can range from 100% (perfect consistency) to 0% (no consistency). If more than two indexers are used, a mean consistency is calculated from each possible pair of indexers.

$$CP\ (\%) = \frac{100A}{A+M+N}$$

RANGE = 0% – 100%

Example: Indexer M uses 11 terms to describe the subjects of an article.
Indexer N uses 10 terms to describe the same article.
There are 8 terms that both indexers used.

A = 8 (the number of terms in agreement)
M = 3 (the number of terms used by M, but not by N)
N = 2 (the number of terms used by N, but not by M)

$$CP\ (\%) = \frac{100 \times 8}{8 + 3 + 2} = \frac{800}{13} = 61.5\%$$

Fig. 1—Hooper's Measure of Indexing Consistency.

The first interindexer consistency study of MEDLINE was performed by Lancaster as part of his evaluation of MEDLARS [5]. Sixteen articles from the data base were each reindexed by three National Library of Medicine (NLM) indexers. Two categories of consistency were measured. One category was checktags plus main headings with subheadings. In this, a main heading/subheading combination was considered not to match the main heading alone (e.g., "HEART/drug effects" does not match "HEART"). The other category was checktags plus main headings only ("HEART/drug effects" matches "HEART"). Consistency percentages as calculated with the Hooper equation were 34.4% and 46.1%, respectively.

Leonard, as part of his investigation into the relationship between indexing consistency and retrieval effectiveness, used ten NLM indexers and a pool of one hundred articles. He also used the same categories that Lancaster measured. The consistency for checktags plus main headings with subheadings was 36.5%, and consistency for checktags plus main headings only was 48.2% [6].

A recent study by Marcetich and Schuyler [7] sought to discover if computer-assisted indexing could improve indexing consistency. Four NLM indexers used routine indexing techniques with fifty English-language articles. Four other NLM indexers indexed these same articles; however, they had access to a list of "suggested" descriptors from *Medical Subject Headings* (MeSH) for each article prepared by AID (Associative Interactive Dictionary), a computer program that suggests possible MeSH terms for articles based on analysis of their abstracts. Although Hooper's equation for consistency was also used, different categories were measured. The indexing consistency for all MeSH terms (excluding checktags and subheadings) was 39% for the routine indexing method and 43% for the computer-assisted method. When printed *Index Medicus* terms only were measured (central concepts of an article), the consistency rose to 48% for the regular method and 55% for the computer-assisted method. A summary of the findings of these articles is given in Table 1.

Several limitations are apparent in these studies. Of necessity, only very small samples of articles were used. The expense of taking indexers away from their regular duties limits the size of these experiments; such small samples, however, may not be representative of the overall consistency of indexing in the data base. Further, because of each study's experimental design, each indexer was

TABLE 1

Summary of Findings from Past Studies

	No. of Articles	Checktags and Main Headings and Subheadings, %	Checktags and Main Headings Only, %	All Main Headings, %		Print Main Headings, %	
				Manual	Computer	Manual	Computer
Lancaster [5]	16	34.4	46.1	NA*	NA	NA	NA
Leonard [3]	100	36.5	48.2	NA	NA	NA	NA
Marcetich and Schuyler [7]	50	NA	NA	39	43	48	55

*NA indicates not applicable.

aware that he was being "tested" on his consistency. This awareness could have produced indexing consistency percentages vastly different from those seen in normal working conditions. Finally, each study measured only a few categories of consistency. MEDLINE searchers use MeSH terminology in many different combinations to limit retrieval. It would be of particular interest to MEDLINE users to know the relative consistency of other categories. The present study used a large sample and normal indexing techniques, and measured nine categories of consistency.

METHODOLOGY

Our sample consisted of journal articles in *Index Medicus* that were indexed twice by NLM. Multiple indexing occurred for one of three reasons. The first is Murphy's law: if anything can go wrong, it will. NLM uses a manual check-in system to route journal issues to their indexing section. Rarely, a duplicate of an already-indexed issue is accidentally sent on for indexing, where it is inadvertently reindexed and reentered into MEDLINE (Lloyd Wommack, personal communication).

The second reason is that a single issue can be published jointly by two journals, and both "issues" then be indexed under each title. For example, a 1976 issue of *Birth Defects* was also published as an issue of *Cytogenetics and Cell Genetics;* the same articles were indexed separately under each journal's title.

The third reason is rare: in two cases of our sample, a journal issue was also selected for indexing as a monograph. An issue of *Monographs in Pathology* (1977; 18) was also indexed as the monograph *The Gastrointestinal Tract;* likewise, an issue of *Current Topics in Molecular Endocrinology* (1974; 1) was also indexed as the monograph *Hormone Binding and Target Cell Activation in the Testis.* Thus, the same articles were indexed twice at different times.

Once identified, these twice-indexed articles are ideal for measuring indexing consistency in MEDLINE, as no artificial testing environment existed at their time of indexing. To find these twice-indexed articles, we scanned the author section of several years of *Index Medicus,* where they appear as duplicate citations, i.e., they are printed twice. Colleagues also assisted by sending us duplicate citations they found in *Index Medicus* and in MEDLINE. A total of 760 articles from forty-two journal issues were identified. Their publication dates ranged from 1974 to 1980. Table 2 gives these twice-indexed journal issues.

All of the citations from each twice-indexed

TABLE 2
SOURCES OF DUPLICATION

Journal issues with duplicate indexing
Acta Eur Fertil 1977 Sep;8(3)
Am J Phys Anthropol 1978 May;48(4)
Am J Surg 1980 Jan;139(1)
Austr Paediatr J 1979 June;15(2)
Aviat Space Environ Med 1976 Sep;47(9)
Bibl Psychiatr 1978;(159)
Birth Defects 1976;12(7)
Br Heart J 1979 Mar;41(3)
Br J Haematol 1979 Nov;43(3)
Bristol Med Chir J 1975 July/Oct;90(335/336)
Circ Shock 1978;5(1)
Clin Gastroenterol 1977 Jan;6(1)
Clin Pediatr 1978 Aug;17(8)
Community Dent Oral Epidemiol 1977 Jan;5(1)
Compr Psychiatry 1979 Jan/Feb;20(1)
Contemp Neurol Ser 1977;15
Contraception 1978 June;17(6)
Cytogenet Cell Genet 1976;16(1/5)
Dan Med Bull 1978 Aug;25(4)
Duodecim 1977;93(19)
Fiziol Zh 1978 Jan/Feb;24(1)
Fiziol Zh 1978 Mar/Apr;24(2)
Folia Parasitol 1978;25(1)
Gynecol Obstet Invest 1978;9(5)
Headache 1978 Jan;17(6)
Hosp Community Psychiatry 1978 Jan;29(1)
Immunochemistry 1977 Jan;14(1)
Isr J Dent Med 1977 Apr;26(2)
J Appl Physiol 1977 Nov;43(5)
J Community Health 1977 Winter;3(2)
J Parasitol 1978 Feb;64(1)
J Reticuloendothel Soc 1978 Jan;23(1)
Jpn J Surg 1978 June;8(2)
Med Klin 1977 Jan;72(1)
Med Klin 1979 Jan 19;74(3)
Monogr Pathol 1977;(18)
Pharmacol Biochem Behav 1976 June;4(6)
Sov Med 1978 Nov;(11)
Ukr Biokhim Zh 1978 Jan/Feb;50(1)
Vopr Kurortol Fizioter Lech Fiz Kult 1978 Mar/Apr;(2)
Vopr Pitan 1978 May/June;(3)
Vrach Delo 1977 Jan;(1)

Monographs with duplicate indexing
Dufau ML, Means AR, eds. Hormone binding and target cell activation in the testis. New York: Plenum Press, 1974.
Yardley JH, et al, eds. The gastrointestinal tract. Baltimore: Williams & Wilkins, 1977.

journal issue were then retrieved from MEDLINE, and the unit records of their entire contents were printed off-line, in the "detailed" format. This format includes the full citation, abstract, complete indexing, and other data.

Using Hooper's equation, we calculated nine categories of indexing consistency from our sample. These categories, their definitions, and our abbreviations are as follows:

1. Checktags (CT): *The Online Services Reference Manual* defines these as "large-volume descriptors routinely checked for in every indexed article" [8]. Examples include HUMAN, ANIMAL, MALE, FEMALE, PREGNANCY, and INFANT.

2. Central concept main headings (*MH): These are MeSH terms that reflect the central concepts of an article. The citation of the article will be printed under these terms in *Index Medicus*. These are also called IM headings, "print" headings, or "starred" headings (the star refers to an asterisk placed before the term in MEDLINE to indicate that it is a central concept). This category is calculated without considering whether subheadings are attached.

3. Main headings (MH): These are any MeSH terms assigned to an article, whether they represent central concepts or the more peripheral points. We included all MeSH terms that were not checktags or geographics, starred or unstarred, without consideration of subheadings. When searching the data base, a MeSH term entered without a central concept indicator will retrieve both major concepts and peripheral points of an article.

4. Central concept main heading/subheading combinations (MH/*SH): When a specific aspect of a subject covered by one of the topical subheadings is considered to be the central concept of an article, the indexer assigns "central concept" status to the combination that is the central concept, not the MeSH term alone.

5. Main heading/subheading combinations (MH/SH): Any main heading with its attached subheading, whether or not it is considered to be a central concept (i.e., starred or unstarred).

6. Subheadings (SH): This category includes any topical subheading attached to any main heading, whether that aspect of the subject covered by the subheading is central or peripheral. These are often called "floating" or "bald" subheadings. Although indexers cannot assign floating subheadings to an article, it is possible to search for a subheading without regard to the main heading to which it is attached.

7. Central concept subheadings (*SH): This category includes those aspects of articles covered by subheadings considered to be the articles' central concepts. We included these subheadings without considering the main headings to which they were attached.

8. Geographics (GEOG): These are category Z terms of MeSH that are assigned to indicate geographic aspects of a subject. They are used routinely for articles on such subjects as epidemiology, legislation, associations, and insurance.

9. Descriptors (DESC): A combination of checktags, main headings, and geographics used in each article, i.e., items 1, 3, and 8.

In addition to measuring these nine categories of consistency, we were interested in whether physical or intellectual factors of a journal article might affect consistency. We noted depth of indexing, indexing priority of a journal, language of the article, length (in pages) of an article, and subject areas of articles with very high and very low indexing consistency.

RESULTS

Categories of Consistency

The mean consistency percentages for each of the nine categories are given in Table 3. The highest consistency was 74.7%, for checktags. A one-tailed t test showed this to be significantly higher ($P < .01$) than the next highest category, central concept main headings at 61.1%. The lowest consistency was 33.8%, for main heading/subheading combinations.

In addition to analyzing individual categories, we also compared those that fell into two natural groupings: central concept and noncentral concept terms, and the use of subheadings, whether

TABLE 3

CONSISTENCY PERCENTAGES FOR EACH CATEGORY

Category	%
Checktags (CT)	74.7
Central-concept main headings (*MH)	61.1
Geographics (GEOG)	56.6
Descriptors (DESC)	55.4
Central-concept subheadings (*SH)	54.9
Subheadings (SH)	48.7
Main headings (MH)	48.2
Central-concept main heading/subheading combinations (MH/*SH)	43.1
Main heading/subheading combinations (MH/SH)	33.8

TABLE 4
NATURAL GROUPINGS OF CATEGORIES OF CONSISTENCY

Category	Consistency Percentage
Central-Concept Terms Compared with Noncentral Concept Terms	
*MH	61.1
MH	48.2
MH/*SH	43.1
MH/SH	33.8
*SH	54.9
SH	48.7
Floating Subheadings Compared with Attached Subheadings	
*SH	54.9
MH/*SH	43.1
SH	48.7
MH/SH	33.8

attached or floating. The consistency percentages of each pair in these groupings were compared and tested using a one-tailed t test. In all cases, the difference between the means was statistically significant ($P < .01$). These comparisons are summarized in Table 4.

Depth of Indexing

Closely related to the choosing of specific terms is the number of terms applied to each article. This is often referred to as the "depth" or "exhaustivity" of indexing. Figure 2 compares the mean number of terms chosen by indexers M and N in our sample. (Geographics are not included in this comparison because no more than one geographic term is typically assigned to any article.) The graph shows the closeness of the number of terms in each category, and t test proved that the differences were not statistically significant.

Journal Indexing Priority

Each of the journals indexed for MEDLINE is assigned an indexing priority: 1, 2, or 3. Priority 1 and 2 journals are more substantive publications and are indexed in greater depth. Priority 1 journals, however, are indexed first, sooner than those at priority 2. Priority 3 journals are indexed with only enough terms to cover the central concepts. We hypothesized that the journal priority might affect the indexing consistency, and calculated the consistency percentage of DESC for each priority: priority 1 (123 articles), 57.5%; priority 2 (432 articles), 53.8%; and priority 3 (205 articles), 57.5%. Although the consistency percentages for each priority are not identical, an analysis of variance (ANOVA) disclosed no statistically significant difference.

Language of the Article

Four languages are included in our sample population of 760 articles: English, Russian, German, and Finnish. It is conceivable that the language of the original article might affect the assignment of indexing terms. We tested this by determining the consistency percentage for DESC for each language. The results are as follows: Russian (216 articles), 58.7%; English (528 articles), 54.5%; German (ten articles), 41.5%; and Finnish (six articles), 41.2%. Again, ANOVA showed no statistically significant difference among the values.

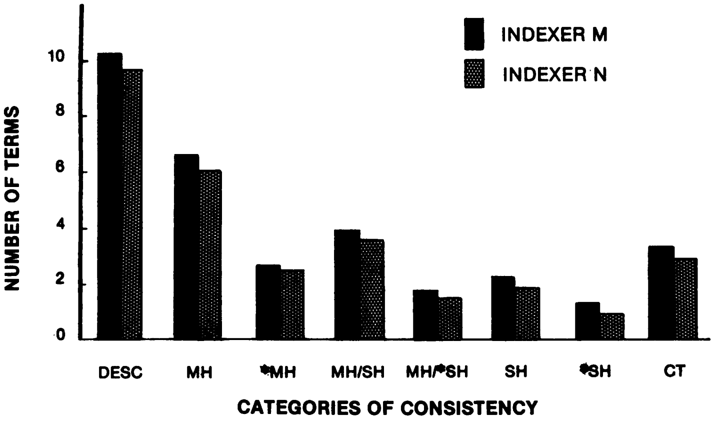

FIG. 2—Comparison of the depth of indexing.

Length of the Article

The articles in our sample were from one to forty-five pages long, with a mean of 6.3 pages. A Pearson correlation showed no statistically significant correlation between length and consistency for any of the nine categories of consistency.

Subject Areas of High and Low Consistency

To determine if there were subject differences between the articles that had high consistency and those of low consistency, we selected two groups of articles based on the consistency percentage in the MH category. The first group contained 110 articles in which the consistency was greater than 70.2% (1 standard deviation [SD] above the mean). The low-consistency group consisted of 129 articles in which consistency was less than 26.1% (1 SD below the mean). We then analyzed the MeSH terms assigned to each article with regard to MeSH Tree Structure Category. When a term appeared in more than one category, we used the one category that seemed most appropriate for that article. A χ^2 test showed a statistically significant difference between the distribution of MeSH terms from the high-consistency articles and that of the low-consistency articles (Figure 3). Category A, B, and D terms appear more often than expected in the high-consistency articles, and category E, F, H, and N terms more often than expected in the low-consistency articles.

Discussion

Although few areas are directly comparable, our results confirm two trends shown in previous studies (Table 1). First, any type of central concept category (*MH, MH/*SH, *SH) has a degree of consistency higher than that of the corresponding noncentral concept category. Second, when subheadings are attached to main headings (MH/*SH, MH/SH), consistency is lowered. Central concepts are the most important aspects in the articles, and are extremely important for subject retrieval both by users of *Index Medicus* and by on-line searchers. As the indexers consistently agree on what constitutes a central concept of an article, searchers can expect a high degree of relevance when using central-concept terms in their search strategies.

Subheadings are used for very specific aspects of a subject. When the indexer attaches a subheading to a main heading this is in essence an additional term assigned to the article, thus increasing the odds against agreement with another indexer. Because the indexing is less consistent when subheadings are attached, the searcher may find less satisfactory results when using main heading/

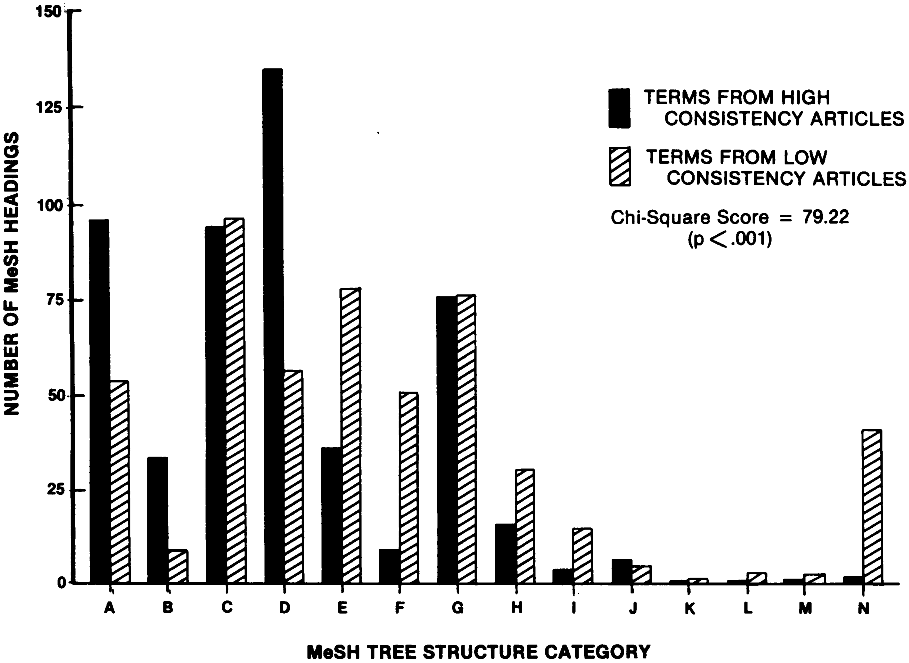

FIG. 3—Comparison of *Medical Subject Headings* Tree Structure categories for high- and low-consistency articles.

subheading combinations. However, on-line computerized data bases allow searching on subheadings, whether starred or unstarred, regardless of the main headings to which they are attached (floating subheadings). Our results indicate that floating subheadings are more consistent than attached subheadings, and so may be used more reliably in search strategies. The searcher must remember, however, that false drops will always occur when floating subheadings are used. If specificity of results is required, then attached subheadings might be a better choice. The trend toward higher consistency in central-concept terms also applies to subheadings: central-concept subheadings, whether attached or floating, show a statistically higher indexing consistency percentage than do the corresponding noncentral-concept subheadings.

The consistency for checktags was the highest (74.7%) of any category we examined. This high percentage probably reflects the fact that there are fewer than fifty checktags, and they are preprinted on the indexing form, which makes them convenient to use. The indexer routinely looks for checktag concepts and can assign any number of them to an article. Because they are used so consistently by the indexers, checktags can be used reliably and effectively by on-line searchers to limit retrieval.

The consistency percentage for geographics (56.6%) may seem surprisingly low: after all, how could an indexer confuse France and Italy? However, the decision facing the indexer is not which geographic term to use, but whether to use one at all. In the articles we studied, inconsistency occurred when one indexer used a geographic and the other did not. These terms can be very useful to limit retrieval to a specific location, but searchers should be aware of this inconsistency and the likelihood of not retrieving relevant citations when geographic terms are used.

Length of the article, language of the article, and journal indexing priority all showed no statistically significant effect on consistency. Furthermore, all nine categories of consistency showed significant agreement among the indexers on the depth of indexing. There seems to be a clear understanding of the number of terms that an article requires. In light of the fact that an average of 250,000 articles are indexed each year for *Index Medicus* and MEDLINE, these four aspects reflect a level of quality control that is exemplary.

Terms from three MeSH Tree Structure Categories appeared more often than expected in the high-consistency articles: category A (Anatomy), category B (Organisms), and category D (Chemicals and Drugs). The high consistency of categories A and B can possibly be explained by the fact that the terminology in these areas is relatively stable, few new terms are added each year, and authors possibly use the older terms more consistently. Unlike categories A and B, many new terms have been added to category D through the years; however, these terms are very specific and allow the indexer little choice. In addition, when the indexer has questions or difficulties with a chemical concept, the article can be referred to the chemical specialist at NLM for clarification and assignment of the correct term.

Terms from four MeSH trees appeared more often than expected in the low-consistency articles: category E (Analytical, Diagnostic, and Therapeutic Technics and Equipment), category F (Psychiatry and Psychology), category H (Physical Sciences), and category N (Health Care). Marcetich and Schuyler also found categories E, H, and N to have lower consistency. They reported that terms from these categories tended to be used to index the research methodology and discussion sections of articles. Similarly, we found that almost one third of the category E terms in our sample were on research methods, primarily in the areas of epidemiologic methods, longitudinal studies, evaluation studies, and clinical trials. This low consistency could be due to the difficulty indexers have in dealing with the researchers' lack of standardized terminology to describe their experiments. This is reflected in a recent editorial from a biomedical research journal that pleaded for standardization in research methods terminology [9]. As the vocabulary of research methodology stabilizes, we expect the indexing consistency in this area to rise.

One can also expect a rise in the consistency with which terms in category F are used. Our study covered articles published between 1974 and 1980, and we found the consistency assignment of category F terms to be significantly lower than expected. Terminology problems in the area of psychiatry and psychology have been pointed out by users in a previous study [10]. However, in 1981 MeSH underwent a great change in category F terminology to correspond with the third edition of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders*. It is appropriate to expect increased consistency in the use of this newer terminology, both by those writing on the subject and by the indexers.

Category N, the terminology of health care, was also an area of low consistency. As Lancaster has pointed out, many factors affect consistency, not the least of which being what he refers to as the "'hardness' or 'softness' of the subject matter being indexed" [11]. Certainly, when compared with anatomical terms, organisms, or chemical names, the terminology of health planning and administration must be considered "soft." Both our study and that of Marcetich and Schuyler revealed the inconsistency of the application of terms in this subject area. Since 1978, major revisions have been made in the vocabulary for health care. The annual revision, deletion, and addition of new terms may eventually overcome the problems in this area. However, until the inherent "softness" of this field "hardens," significant improvement in indexing consistency probably will not be seen.

Conclusions

What do our findings mean? Obviously, we wanted to point out ways in which searchers can improve their results. Also, we noted areas where NLM can improve. But do the present figures indicate relatively good indexing or relatively poor indexing? Unfortunately, consistency in MEDLINE cannot be compared with consistency in any other bibliographic data base in the area of biology and medicine. If indexing consistency has been studied in other biomedical data bases, the results have not been published.

How, then, can one compare the performance level of indexing consistency in MEDLINE? Several studies have examined the disagreements among physicians over clinical findings, diagnoses, and management decisions. Certainly, one can see situations similar to what indexers face. For example, "when three cardiologists interviewed 57 men with chest pain, 54% of the men were judged by at least one clinician to have a history compatible with angina pectoris. However, in only 30% of the 57 patients did all three clinicians agree about the history, and if one of the cardiologists concluded that a given patient had angina pectoris, the other two agreed with him only 55% of the time" [12].

It seems that, realistically, the upper level of consistency for any type of intellectual choice falls far below 100%. We believe that MEDLINE, with its excellent controlled vocabulary, exemplary quality control, and cadre of highly trained indexers, probably represents the state of the art in manually indexed data bases.

References

1. Leonard LE. Inter-indexer consistency studies 1954–1975: a review of the literature and summary of study results. Champaign, Ill.: University of Illi-Graduate School of Library Science, 1977 (Occasional Papers, no. 131).
2. Cooper WS. Is inter-indexer consistency a hobgoblin? Am Doc 1969 July;21(3):268–78.
3. Leonard LE. Inter-indexer consistency and retrieval effectiveness: measurement of relationships, Thesis. Champaign, Ill.: University of Illinois, 1975.
4. Hooper RS. Indexer consistency tests: origin, measurement, results, and utilization. Bethesda, Md.: IBM Corporation, 1965 (TR95-56).
5. Lancaster FW. Evaluation of the MEDLARS demand search service. Washington, D.C.: Na-Library of Medicine, 1968.
6. Leonard LE. Op. cit.: 1975, p. 124.
7. Marcetich J, Schuyler P. The use of AID to promote indexing consistency at the National Library of Medicine. Paper presented at the Eighty-first Annual Meeting of the Medical Library Association, Montreal, Quebec, Canada, June 1981.
8. Online services reference manual. Bethesda, Md.: National Library of Medicine, 1982:134.
9. Meinart CL. Terminology: a plea for standardization. Controlled Clin Trials 1980 Sep;1(2):97–9.
10. McCarthy SE, Maccabee SS, Feng CC. Evaluation of MEDLINE service by user survey. Bull Med Libr Assoc 1974 Oct;62(4):367–73.
11. Lancaster FW. Op. cit.: 1968, p. 179.
12. Sackett DL. Clinical disagreement: how often it occurs and why. Can Med Assoc J 1980 Sep 20;123:499–504.

Received September 1982; revision accepted November 1982.